

Section 1.2

Check Your Understanding, page 29:

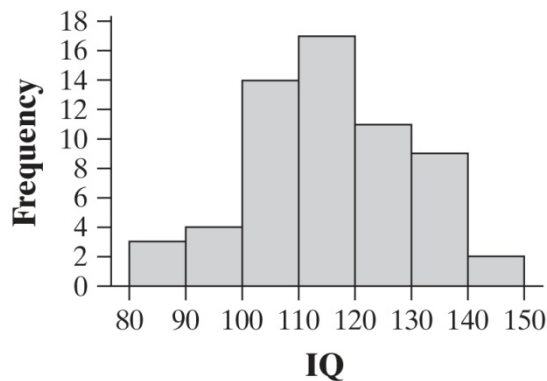
1. This distribution is skewed to the right and unimodal.
2. The midpoint of the 28 values is between 1 and 2. So a typical student in the sample has 1 or 2 siblings.
3. The number of siblings varies from 0 to 6.
4. There are two potential outliers: those students reporting 5 and 6 siblings.

Check Your Understanding, page 32:

1. Both males and females have distributions that are skewed to the right, though the distribution for the males is more heavily skewed. It appears that females typically have more pairs of shoes than males. The midpoint for the males was 9 pairs while the midpoint for the females was 26. The number of shoes owned by females varies more (from 13 to 57) than the number of shoes owned by males (from 4 to 38). Finally, the male distribution has three likely outliers at 22, 35, and 38. The females do not have any likely outliers.
2. b
3. e
4. c

Check Your Understanding, page 38:

1. One possible histogram is below:



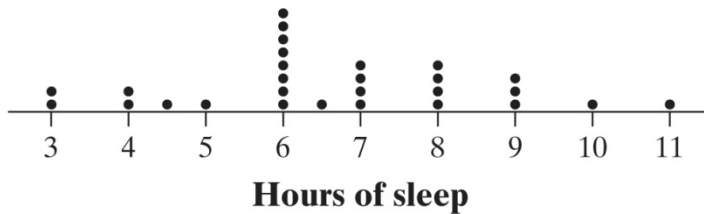
2. The distribution is roughly symmetric and bell-shaped. The typical IQ appears to be between 110 and 120 and the IQs vary from 80 to 150. There do not appear to be any outliers.

Check Your Understanding, page 39:

1. This is a bar graph because field of study is a categorical variable.
2. It would not be correct to describe this graph as right-skewed because the variable is categorical and the categories could be listed in any order on the horizontal axis.

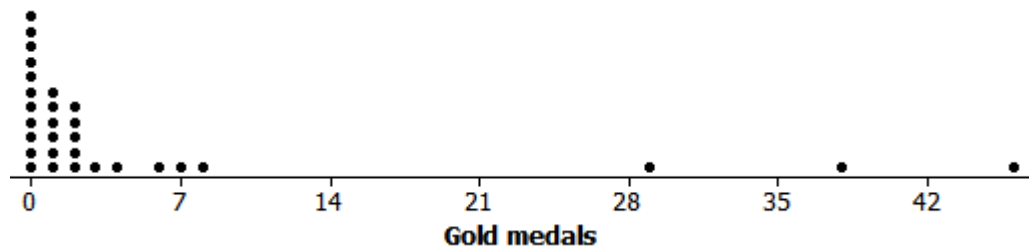
Exercises, page 41:

1.37 (a) The graph is shown below:



(b) The distribution is roughly symmetric with a midpoint of 6 hours. The hours of sleep vary from 3 to 11. There do not appear to be any outliers.

1.38 (a) The graph is shown below.



The distribution is strongly skewed to the right with a peak at 0, which indicates that many countries did not win any gold medals. The midpoint is 1 gold medal and the number of gold medals varies from 0 to 46. There are outliers at 29 (Great Britain), 38 (China), and 46 (United States).

(b) No, this does not seem to be a representative sample since 19 out of the 30 countries in the sample (or about 63%) won gold medals. Overall, only about $54/205 = 26\%$ won gold medals.

1.39 (a) The dot above -1 represents a game where the opposing team won by 1 goal.

(b) All but 4 of the 25 values are positive, which indicates that the U.S. women's soccer team had a very good season. They won $21/25 = 84\%$ of their games, tied the other team in $3/25 = 12\%$ of their games, and only lost $1/25 = 4\%$ of their games.

1.40 (a) The dot above 12 represents a car that got 12 mpg more on the highway than it did in city driving.

(b) From the dotplot we see that the EPA mpg rating is higher on the highway than in the city for all but one of the cars. These other 23 cars all averaged at least 7 mpg more on the highway than in the city.

1.41 As coins get older they get taken out of circulation and new coins are introduced. So most coins in someone's pocket will be from recent years, but there may be a few from previous years.

1.42 The shape of this distribution is roughly symmetric and somewhat uniform. This isn't surprising because we expect that the last digit in a phone number is equally likely to be any digit from 0 to 9.

1.43 Both distributions are roughly symmetric and have about the same amount of variability. The center of the internal distribution is greater than the center of the external distribution, indicating that external rewards do not promote creativity. Neither distribution appears to have outliers.

1.44 The distributions of sugar content on shelves 1 and 3 are both skewed to the right while the distribution of sugar content on shelf 2 is skewed to the left. The center for the distribution of sugar on shelf 2 is higher than the center for the distributions of sugar on shelves 1 or 3, partially supporting the critics' claim that sugary cereals are where kids can see them. All three distributions have about the same amount of variability and none of the distributions appear to have any outliers.

1.45 (a) If we had not split the stems, most of the data would appear on just a few stems making it hard to identify the shape of the distribution.

(b) Key: 12 | 1 means that 12.1% of that state's residents are aged 25 to 34.

(c) The distribution of percent of residents aged 25-34 is roughly symmetric with a possible outlier at 16.0%. The center (midpoint) is around 13%. Other than the outlier at 16.0%, the values vary from 11.4% to 15.1%.

1.46 (a) If we had not split the stems, most of the data would appear on just a few stems making it hard to identify the shape of the distribution.

(b) Key: 2|3 means that an 8-ounce serving of that soft drink has 23 mg of caffeine.

(c) This distribution is slightly skewed to the right. The center is around 28 mg and the values vary from 15 mg to 47 mg. There do not appear to be any outliers. Because all of the values are below 48 mg, all of these drinks meet the USFDA's limit.

1.47 (a) The stemplots are given below:

Without splitting stems

```
6|0 3 5 5 7
7|0 1 2 4 4 8 8 9 9 9
8|1 1 3 6 6 7
9|0 6
```

With splitting stems

```
6|0 3
6|5 5 7
7|0 1 2 4 4
7|8 8 9 9 9
8|1 1 3
8|6 6 7
9|0
9|6
```

Key: 6 3 = 630 mm of rain

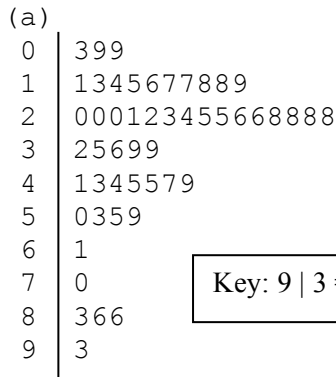
The stemplot with split stems makes it easier to see the shape of the distribution.

(b) The distribution is slightly skewed to the right with a center near 780 mm, and values that vary from around 600 mm to around 960 mm. There do not appear to be any outliers.

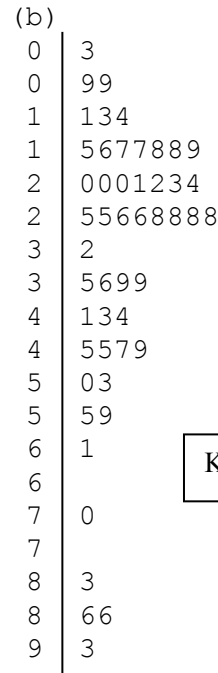
(c) Monsoon rainfall was below 850 in 18 of the 23 El Niño years, and only exceeded 900 mm in one of those years. In El Niño years there is typically less rain than in other years.

1.48 (a) and (b) The stemplots are shown below. The stemplot with the split stems shows the skewness, gaps, and outliers more clearly.

(c) The distribution of the amount of money spent by shoppers at this supermarket is skewed to the right, with a center around \$28, and values that vary from \$3 to \$93. There are a few gaps (from \$62 to \$69 and \$71 to \$82) and some outliers on the high end (\$83, \$86, and \$93).



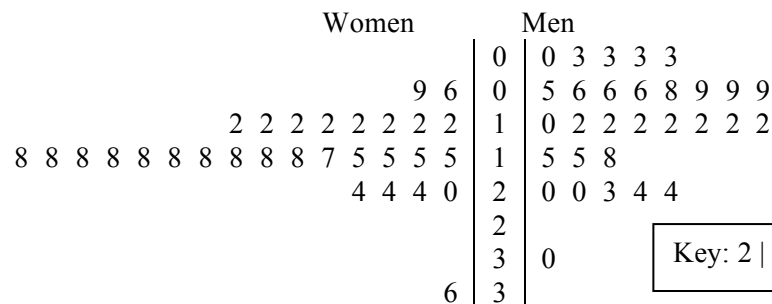
Key: 9 | 3 = \$93



Key: 9 | 3 = \$93

1.49 (a) Not only are most responses multiples of 10; many are multiples of 30 and 60. Most people will round their answers to the nearest hour or half-an-hour. The students who claimed 360 minutes (about 6 hours) and 300 minutes (about 5 hours) of studying on a typical weeknight may have been exaggerating.

(b) The stemplots suggest that women (claim to) study more than men. The center for women (about 175 minutes) is greater than the center for men (about 120 minutes).



Key: 2 | 3 = 230 minutes

1.50 (a) The stemplot is given below:

Division I-AAA		Division V-AA
	8	3
	7 7 6 1	4 4 5 6
8 6 5 4 3 2 1	5	4 8
8 7 5 5 4 4 3 3 2 2	6	0 2 6 6 7 7 9
8 7 6 4 1 1	7	2 4
	7	8
	1	9
	6	10

Key: 9 3 = 93 points

(b) Both distributions are roughly symmetric, although the distribution in Division I-AAA is unimodal and the distribution in Division V-AA has more than one peak. While the Division V-AA center (midpoint = 66) is slightly larger than the Division I-AAA center (midpoint = 63.5), the scores in both divisions are about equally variable, with values that vary from 38 to 106 in Division I-AAA and values that vary from 36 to 98 in Division V-AA. The team that scored 106 points in Division I-AAA might be an outlier.

1.51 (a) The distribution is slightly skewed to the left and unimodal.

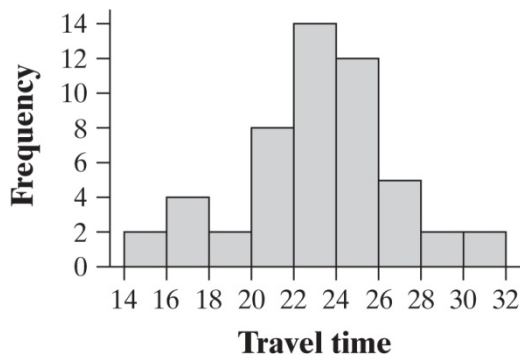
(b) The center is between 0% and 2.5%.

(c) The highest return was between 10% and 12.5%. Ignoring the low outliers, the lowest return was between -12.5% and -10%.

(d) About 37% of these months (102 out of 273) had negative returns.

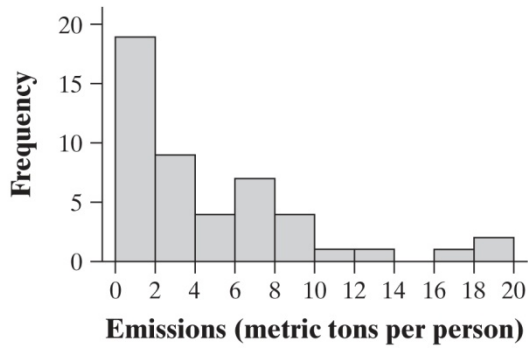
1.52 The distribution of lengths of words in Shakespeare's plays is skewed to the right and unimodal. The center is around 4 letters. The lengths of the words vary from 1 to 12 letters.

1.53 (a) The graph is given below:



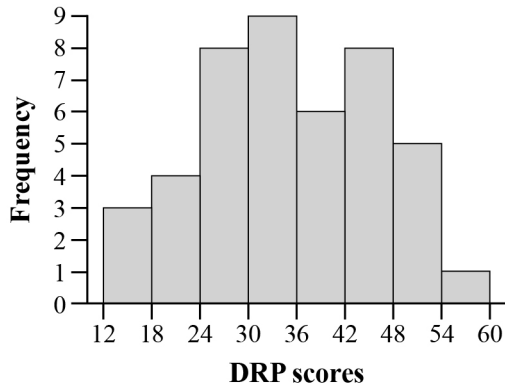
(b) The distribution of travel times is roughly symmetric. The center is near 23 minutes and the values vary from 15.5 to 30.9 minutes. There does not appear to be any outliers.

1.54 (a) The graph is given below:



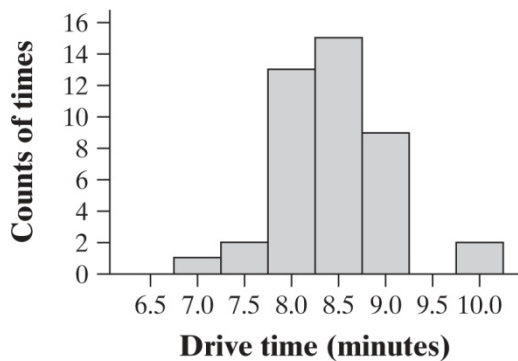
(b) The distribution of the emissions is skewed to the right with center near 4 metric tons per person. The values of CO₂ vary from 0.1 to 19.6 metric tons per person. There appear to be three outliers: Canada, Australia, and the United States.

1.55 The graph is given below:



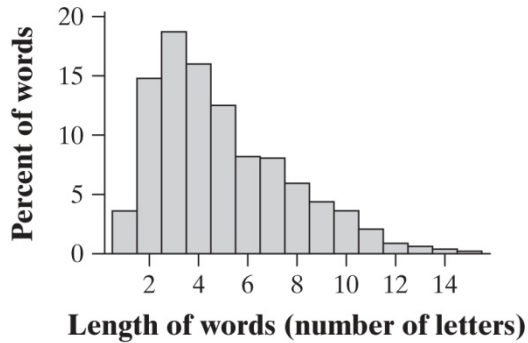
The distribution of DRP scores is roughly symmetric with center around 35. The DRP scores vary from 14 to 54. There doesn't appear to be any outliers.

1.56 The graph is given below:



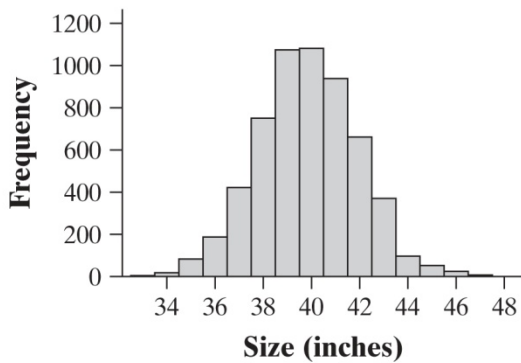
This distribution of drive times is roughly symmetric with a center around 8.5 minutes and values that vary from 6.75 to 10.17 minutes. There are no clear outliers, although times around 10 minutes might be considered outliers.

1.57 (a) The histogram below shows that the distribution of word lengths is skewed to the right and single-peaked. The center is around 4 letters, with words that vary from 1 to 15 letters. There doesn't appear to be any outliers.



(b) There are more 2, 3, and 4 letter words in Shakespeare's plays and more very long words in *Popular Science* articles.

1.58 (a) The histogram is given below:

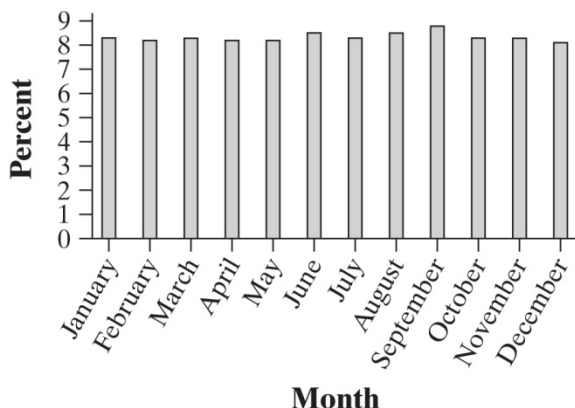


(b) The distribution of chest sizes is roughly symmetric with center around 40 inches and values that vary from 33 inches to 48 inches. This information is important so that the military can order the correct distribution of uniform sizes.

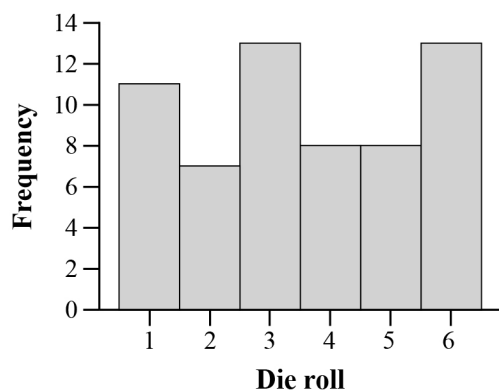
1.59 It is difficult to effectively compare the salaries of the two teams with these two histograms because the scale on the horizontal axis is very different from one graph to the other.

1.60 Both distributions are skewed to the right, but the distribution of salaries for the Yankees has a longer right tail. The salaries for the Yankees have a higher center and more variability. Finally, the Yankees have an outlier somewhere between \$32,000,000 and \$36,000,000.

1.61 A bar graph should be used because birth month is a categorical variable. A possible bar graph is given below:

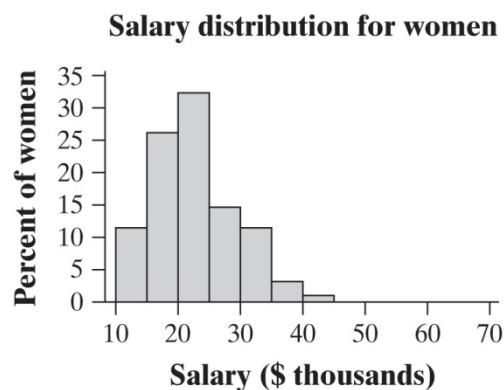
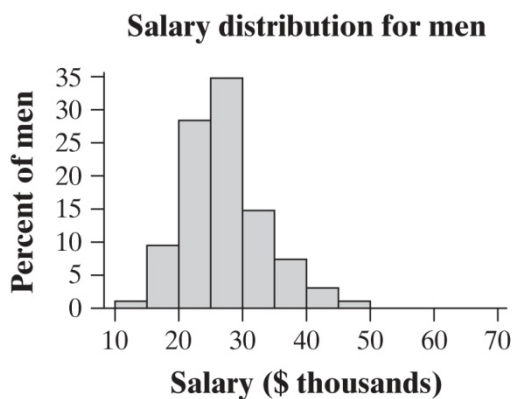


1.62 A histogram should be used because die roll is a quantitative variable. A possible histogram is given below:



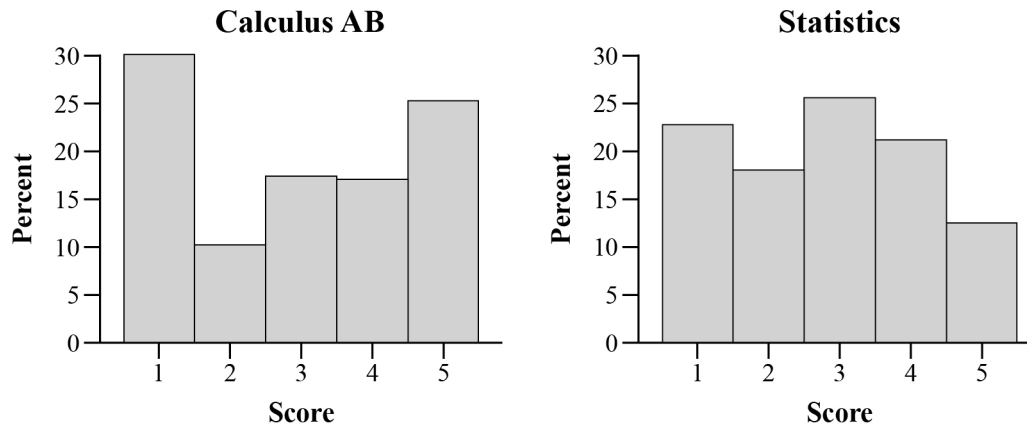
1.63 (a) The percents for women sum to 100.1% due to rounding errors.

(b) Relative frequency histograms are shown below because there are considerably more men than women.



(c) Both histograms are skewed to the right, with the women's salaries generally lower than the men's. The midpoint for women is the interval from \$20,000 to \$25,000, and the midpoint for men is the interval from \$25,000 to \$30,000. The distributions of salaries are about equally variable and the table shows that there are some outliers in each distribution who make between \$65,000 and \$70,000.

1.64 (a) Relative frequency histograms are shown below because there are more calculus than statistics scores.



(b) The shapes of the two distributions are very different. The distribution of scores on the AP Calculus AB exam has a peak at 1 and another slightly lower peak at 5. The distribution of scores on the AP Statistics exam, however, is more uniform with scores of 1 and 3 being the most frequent and scores of 5 being the least frequent. The center of both distributions is 3. Although scores on both exams vary from 1 to 5, there are more scores close to the center on the AP Statistics exam and more scores at the extremes on the AP Calculus exam.

1.65 The distribution of age is skewed to the right for both males and females. Among the younger Vietnamese, there are more males than females. After age 35, however, females seem to outnumber the males. This means that the center of the female distribution is likely to be a little greater than the center of the male distribution. Both distributions have about the same amount of variability. Finally, most Vietnamese are young with a peak frequency for those in their 20's. After age 30, the numbers start to decline fairly rapidly.

1.66 The distribution of age is slightly skewed to the left for both males and females. The prediction for China in 2050 suggests that each age group until age 65 will have slightly more men than women. After age 65, it appears that each age group will have slightly more women than men. This means that the center of the female distribution is likely to be a little greater than the center of the male distribution. Both distributions have about the same amount of variability. Finally, the largest group of Chinese will be in their late 50s and early 60s in 2050.

- 1.67 (a) This graph represents the amount of studying. We would expect some students to study very little, but many more students to study a moderate amount. Any outliers would likely be high outliers, leading to a right-skewed distribution.
- (b) This graph represents the right vs. left-handed variable. About 90% of the population is right-handed and because 0 represents right-handed people we would expect a much higher bar at 0 than at 1.
- (c) This graph represents the gender of the students. We would expect a more similar percentage of males and females than we would for the right-handed and left-handed students.
- (d) This graph represents the heights of the students. We expect many heights near the average and a few very short or very tall people.

- 1.68 (a) Radio station is a categorical variable, so use a bar graph – one bar for each station.
- (b) Because hours studied per week is quantitative, either use a dotplot, stemplot, or a histogram.
- (c) Because calories consumed is quantitative, either use a dotplot, stemplot, or a histogram.

1.69 a

1.70 a

1.71 c

1.72 e

1.73 d

1.74 b

- 1.75 (a) The individuals are Major League Baseball players who were on the roster on opening day of the 2012 season.
- (b) There are six variables besides name. Two of them are categorical (team, position) and the other 4 are quantitative (age, height, weight, and salary).

- 1.76 (a) Generally, more people “love” the newer devices such as the iPod, Broadband, and HDTV. Those less “loved” are older technologies like cable TV and pay TV.
- (b) It would not be appropriate to make a pie chart with these data because the categories are not dividing up a whole into pieces. Individuals could be represented in more than one bar.

- 1.77 (a) There were $10 + 9 + 24 + 61 + 206 + 548 = 858$ observations in all. Of those, $10 + 61 = 71$ were elite players. So $71/858 = 8.3\%$ were elite soccer players. There were $10 + 9 + 24 = 43$ who had arthritis, which means that $43/858 = 5.0\%$ of the people had arthritis.
- (b) 10 of the 71 elite players had arthritis. This means that $10/71 = 14.1\%$ had arthritis. 10 of the 43 people who had arthritis were elite players. This means that $10/43 = 23.3\%$ of those with arthritis were elite soccer players.

1.78 The percent of each group who have arthritis is $10/71 = 14.1\%$ for the elite soccer players, $9/215 = 4.2\%$ for the non-elite soccer players and $24/572 = 4.2\%$ for the people who did not play. This suggests an association between playing elite soccer and developing arthritis. This can also be seen in the following bar graph:

